The Street Lamp Paradox : Analysing Students' Evaluation of Teaching through Qualitative and Quantitative Approaches

Ghislandi, Patrizia

2020-06

Ghislandi , P , Raffaghelli , J , Sangrà , A & Ritella , G 2020 , ' The Street Lamp Paradox : Analysing Students' Evaluation of Teaching through Qualitative and Quantitative Approaches ' , Journal of Educational, Cultural and Psychological Studies , no. 21 , pp. 65-85 . https://doi.org/10.7358/ecps-2020-021-ghis

http://hdl.handle.net/10138/325715
https://doi.org/10.7358/ecps-2020-021-ghis

Educational, Cultural and Psychological Studies
JOURNAL of

21
June 2020

Parte Monografica
E-learning per l'istruzione superiore: nuove indagini empiriche

Monographic Section
E-learning for higher education: new empirical investigations

The Street Lamp Paradox: Analysing Students' Evaluation of Teaching through Qualitative and Quantitative Approaches

Patrizia Ghislandi [1] - **Juliana Raffaghelli** [2]
Albert Sangrà [2] - **Giuseppe Ritella** [3]

[1] *Università di Trento - Department of Psychology and Cognitive Science (Italy)*

[2] *Universitat Oberta de Catalunya (UOC) - Department of Psychology and Educational Sciences (Spain)*

[3] *University of Helsinki - Department of Education (Finland)*

DOI: https://dx.doi.org/10.7358/ecps-2020-021-ghis

patrizia.ghislandi@unitn.it
jraffaghelli@uoc.edu
asangra@uoc.edu
giuseppe.ritella@helsinki.fi

IL PARADOSSO DEL LAMPIONE: ANALIZZARE, ATTRAVERSO APPROCCI QUALITATIVI E QUANTITATIVI, LA VALUTAZIONE DI UN INSEGNAMENTO ACCADEMICO DA PARTE DEGLI STUDENTI

Abstract

This paper presents a study about Students Evaluation of Teaching (SET). Student surveys delivered at the end of the academic course is an approach currently applied in all Italian university courses. However, the quality of teaching is a multi-layered phenomenon whose evaluation might require additional methods. In this study, concerning a course offered during the academic year 2016/2017, we propose the qualitative analysis of students' final essays, which includes the students' opinion about teaching quality. The qualitative evaluation of teaching is then compared with the results of the standard SET survey. 48 students filled in the survey, while 47 delivered their final essay (a corpus of

about 650 text pages). Our study shows that while the standard survey provides an overall picture useful at institutional level for Quality Assurance (QA), the qualitative approach captures an accurate account of students' reactions, sensitive to the pedagogical approach adopted. This provides additional information on the students' perspectives regarding the specific features of the course. Our case study suggests that the integration of traditional SET survey with qualitative teaching evaluation approaches, at least for innovative courses based on socio-constructivist learning, might provide information, overlooked in the Italian SET survey, that is useful for Quality Enhancement (QE) of teaching.

Keywords: Qualitative evaluation of teaching; Quality assurance; Quality enhancement; Socio-constructivist learning; Student evaluation of teaching.

1. Introduction

Evaluating the quality of academic teaching is a challenging problem. The impressive literature about the problem of quality of teaching and learning in Higher Education (HE) demonstrates the complexity of this construct (Henard & Roseveare, 2012; Communique, 2015). Teaching quality can be considered a multi-layered (i.e. implemented at several levels) and a multi-perspective (i.e. with several stakeholders to be engaged in the process of evaluation) phenomenon (Ehlers, 2004; Cole *et al.*, 2004; Mahoney, 2012; Yang, 2015). In the light of customer satisfaction approaches, the evaluation of teaching quality by students, or Students Evaluation of Teaching (SET), has been implemented since the 70's as a widespread approach to capture the perspective of those who receive teaching as a service. However, the question concerning which are the best ways of implementing this form of evaluation is still open. Specifically, it remains unclear how students' perspective can be captured in a way that is fruitful for the enhancement of teaching and learning. Students' perspective is complex and includes multiple elements that have great potential in supporting teachers' awareness of quality issues in their activity. However, the methods, procedures and metrics currently adopted encompass several pitfalls.

Although quantitative SET is dominant because of its efficiency in collecting and analyzing data, its focus could be limited and of little help to get data that can effectively promote significant enhancement of quality teaching. This applies particularly to courses offering teaching innovations which exceed students' expectations, and to courses with socio-constructivist design, which entail more student cognitive workload (Jonassen, 1992). In

this regard, qualitative approaches have been promoted to study the peculiarities of socio-cultural aspects and local situations requiring an idiographic analysis that, through an inductive process, make sense of emerging practices or phenomena (Guba & Lincoln, 1989; Patton, 1990; Shaw, 1999).

The street lamp paradox of the title alludes to a tale, told few lines hereafter, attempting to introduce, through a brief parable, the difficulties faced while evaluating innovative teaching practices through standardized methods.

A policeman saw a drunkard looking for his keys under a street lamp light. «What are you looking for here, good fellow?» said the policeman. «I am looking for my keys […] 'cause I lost them […]». «You lost them here?». «I don't know […]. But here is where the light is […]».

The story is not new, as well as the streetlight effect, already cited by Kaplan (1973), and reported as the «principle of drunk research».

The history of quality evaluation has always been an attempt of reducing complexity to something simpler that can be measured. But, as the drunkard story teaches us, while a research could be simpler, the most significant data not always come from what can be easily measured, entailing the risk of overlooking what is relevant.

2. The context of this study

In the academic year 2016/17, 49 students of an undergraduate program on Interfaces and Communication Technologies at an Italian university, attended a blended course (Sangrà *et al.*, 2012) on eLearning design.

The eLearning Design (eLD) course was based on authentic tasks and innovative ways of assessment (including self and peer assessment). First, the students formed 10 groups of 4-5 people. In each group there were 5 rotating roles: the president, the moderator (moderating the forum debates), the editor (in charge of timely delivery of homework assignment), the group log's writer, the snitch (who could access all the other groups' documents and forums). For each of the eight modules group homework was planned, as well as timely feedback. After some theoretical modules about eLearning design, each group was invited to design, create, and deliver a short eLearning course, as well as to attend the course created by the so-called «tandem group». This allowed to implement an assessment approach that integrated self and peer evaluation, which has proven to be an effective tool for assessment (Topping, 1998; Falchikov, 2005; Nicol *et al.*, 2014; Grion & Tino, 2018). In addition, *ad hoc* rubrics were adopted

(Ghislandi, 2012; Ghislandi *et al.*, 2012). The course ended with group presentations followed by a question-and-answer session. It must be highlighted that the active methods are less frequent in university teaching, but they are yet important to trigger deep forms of learning and competence (Laurillard, 1993; Allendoerfer *et al.*, 2016).

Finally, the students prepared an essay, an exercise of meta-reflection on their whole learning path. Awareness of process of learning is one of the important learning objectives of the eLD course and, as Flavell (1979) say, a critical factor to successful learning. Moreover, as claimed by Gibbs (2010), «one of the most telling indicators of the quality of educational outcomes is the work students submit for assessment, such as their final year project or dissertation. These samples of students' work are often archived, but rarely studied».

The final essay was also used by the teacher to collect qualitative data (provided in free text) about the students' evaluation of teaching.

Before the final assessment, the attending students were requested to anonymously fill out a survey collecting Student Evaluation of Teaching (SET).

3. Related studies

In this section, the main background constructs are introduced, providing a conceptual basis for our study.

Quality of teaching is embedded within the classical discussion about quality in education, based on a huge literature, where quality is purported as a complex, multi-layered and multi-perspective problem (Harvey & Green, 1993; Harvey, 2004). In this regard, there are two quality evaluation types, Quality Assurance (QA) and Quality Enhancement (QE), as many authors claimed (Raban, 2007; Elassy, 2015). The latter concept entails an idea of a process and a continuum, which is not present in the former. For Raban (2007), although QA can assure a standard level of quality, it is not a perfect match with the quality enhancement of teaching: «What is wrong with the conventional architectural style of our quality management systems? […] they are unfit for their declared purpose of securing significant improvements in the quality of teaching». Moreover, for Elassy (2015): «the concepts of QA and QE should be dealt as part of a continuum and showed the need for both as an ongoing process in HE institutions».

SET has a more specific focus on the study of teaching quality, which begins in the Anglo-Saxon countries, in the '70s and '80s of the last cen-

tury, mainly achieved with quantitative surveys, as the scientific literature about the same topic say. The SET movement has many proponents and opponents, and we can find a documented analysis of the different position in a recent study (Uttl *et al.*, 2017). The proponents claim that SET is cheap and convenient and a mean to serve public accountability, a concept strictly related to Quality Assurance.

Lattuca and Domagal-Goldman (2007), and Ory (2001) point out that while students are effective judges when evaluating the teacher's clarity of exposure and organization of contents, as well as the ability to deliver content and to facilitate interactions, they are instead non-trustable when judging the actual contents, because other stakeholders should evaluate what is necessary to achieve adequate preparation. In a recent and enlightening systematic literature review of research published after 2000 (Spooren *et al.*, 2013), it is clarified that researchers tend to equate students teaching evaluation with teaching quality. That is, we tend to conflate students' opinion about the teaching and the knowledge students achieve, and this assumption is only partially true (Buck, 1998; Uttl *et al.*, 2017). However, many students' valid ideas and suggestions remain unused, because teachers that do not perceive SET instruments as valid tend to ignore them (Rienties, 2014).

In the Italian context, where our study is located, we can find some early studies about the evaluation of pedagogical innovations (Giovannini, 1988; Losito, 1996). More specifically SET has been studied from the end of the last century, and the available literature tends to take for granted the use of surveys as a means of collecting students' data for evaluation of teaching's quality. A report on students' evaluation of teaching, containing a survey proposal, was published in 1998 (Bernardi *et al.*, 1998). Many other studies followed later on (Fabbris, 2002; Pagani & Seghieri, 2002; Rampichini *et al.*, 2004; Chiandotto *et al.*, 2005) mainly addressing data collection issues from the statistical and organizational point of view. From 2013, the system proposed by the National Agency of University System Evaluation, or Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR), regarding academic evaluation and accreditation, has been operating with the goal to improve teaching and research quality, and also collecting students' opinions. In Bertaccini (2015) we can find a documented history of the Italian teaching quality evaluation, as well as a critical analysis of some of the adopted solutions. Nowadays, the literature offers papers proposing new SET surveys (Bertaccini *et al.*, 2019) while more pedagogical studies are also starting to emerge (Lalla, 2006; Braga *et al.*, 2014; Giovannini & Silva, 2014; ANVUR, 2018; Balzaretti & Vannini, 2018).

As it comes out of the debate on quantitative and qualitative research methods (Lincoln *et al.*, 2011), in the mentioned studies there is an emerging concern with survey as a data collection technique that encompasses measurements and statistical inference, entailing the assumption of objectivity, but whose coverage of the phenomena is limited. Instead, qualitative studies, while limiting the scale of sampling, could shed light over peculiarities, local cases and, above all, examine the impact of emergent practices along a continuum of quality enhancement (Ghislandi *et al.*, 2013). In a nutshell, it is important to understand when synthesis properly and validly represents a cultural and social phenomenon concerning the teaching/learning processes, and when it is necessary to conduct further analysis.

As Macdonald (2006) claims,

> It is worth remembering that there is no single right way of achieving our educational aims nor of evaluating them. At times we need to reflect on whether a more divergent approach might yield more useful results than the tried and tested convergent methods that everyone else seems to use.

4. Methodology

The research question the authors held as a reference during their study is the following: *Which forms of SET are effective in evaluating non-traditional and emergent teaching practices?*

This research is based on a case study within which mixed methods were used to investigate students' evaluation of teaching. The mixed methods design adopted in this case was the triangulation design, mixing methods into a multilevel model, or QUAN-QUAL design (Creswell & Plano Clark, 2007). To answer the research question formulated above, the data collection was based on the analysis of two methods of SET: the standard survey and a final essay.

The standard survey about SET relates to the Italian higher education questionnaire delivered, at a national level, to all the students and composed of 12 questions. It is centered on students' opinion concerning teaching. The standard survey preserves students' privacy and anonymity. Students answer using an ordinal, four-level Likert scale («definitely not», «rather no», «rather yes», «definitely yes» and the option «not foreseen»). In our research the responses to the survey's questions of the students attending the eLD course were used as primary data for the quantitative analysis.

An initial version of this questionnaire was developed about two decades ago and the whole process of teaching evaluation based on the

survey has undergone several iterations of improvement and research across the years (Bertaccini, 2015). Nevertheless, it has been highlighted that some items of the survey might raise some interpretive doubts (Bertaccini, 2015). Related to such interpretive problems, we argue that some of the items seem to imply the primacy of a transmissive approach and might lead to misinterpretation when applied to courses that are not based on such an approach. For example, the Q8 (*Tab. 1*) states «Are the integrative teaching activities (tutorials, laboratories, etc.), where existing, useful to learn the subject?». This question seems to imply that lecturing is considered as the main teaching activity, while practical activities such as laboratories, where students play an active role, are considered secondary. Moreover, the focus on teachers' clarity in explaining topics and on the students «understanding» of a subject (which overlooks skills and competences) support our interpretation. Indeed, some innovative pedagogical approaches such as knowledge building (Cacciamani & Messina, 2011), the dialogical pedagogy (Wegerif, 2006), inquiry learning (Kuhn *et al.*, 2000); object-based learning (Muukkonen *et al.*, 2011), collaborative and constructivist participation model (Loperfido *et al.*, 2011) – just to make a few examples – give primacy to the orchestration of students practical and collaborative activities instead of direct instruction for the development of disciplinary and interdisciplinary competences, in addition to content knowledge. Within these frameworks, usually based on a socio-constructivist framework, the activities that in the survey are considered as «integrative activities», are the core of the pedagogical intervention. There are also some preliminary researches showing how socio-constructivist approaches might contribute to an effective development and transfer of skill and competences useful for the students' working life (Ritella *et al.*, 2020). This preliminary analysis allowed us to develop the research question of this research, which is based on the argument that the current survey is not able to make visible important aspects of constructivist learning. Therefore, we designed the research discussed in the present article to explore through the final essays what are the dimensions of a constructivist course that are considered as the most meaningful for the students.

With the rationale presented above the attending students of the eLearning Design course produced a non-anonymous final essay about their project group work. The second part of the essay was devoted to the reflections regarding the teaching quality of the course: learning design, learning environment and resources deployment, online learning and teaching as experience, teamwork, personal portfolio, feedback, concept maps, rubrics. Therefore, the students were given the possibility to reflect on their learning processes and to evaluate the teaching approach.

During the triangulation, data is merged and interpreted. In our research, the two methods were then compared to investigate how each of them was able to capture the students' opinions on their learning experience. As stated in the literature, quantitative and qualitative findings cannot be compared directly, but the structural characteristics of the interpretations made could be explored and discussed (Creswell & Plano Clark, 2011).

The data were elaborated as follows:

1. Descriptive statistics were applied to the survey. The questionnaire completed by the students yielded data used to generate univariate descriptive statistics counting the frequencies of answers. Also, percentages were calculated on the basis of aggregated data for each item of the questionnaire.
2. Thematic analysis on the final essay. Thematic analysis is an approach commonly applied to corpus of text as a qualitative technique of analysis. It consists on identifying, analyzing and interpreting patterns of meaning (or «themes») within qualitative data. The essays were treated as a corpus of analysis. Pseudonyms were used so that the identity of the participants could not be recognized. Two researchers made an independent tentative proposal of codes, which was followed by an inter-codification process and the discussion of categories based on 1171 coded excerpts. A third researcher, which worked on a blind basis, coded the 9% of the corpus. The Kappa obtained was 0.84, thus a nearly perfect agreement (Landis & Koch, 1977). The instruments adopted for the process of codification were quite simple: PDFs sent by the students were read and codified and the relevant units of text were collected in a table.
3. Once the procedure of thematic analysis was concluded, content analysis was applied. This consists of counting codes to synthesize the thematic findings. The descriptive statistics obtained are not meant for inferential analysis, but serve the purpose of identification and visualization of emergent themes (Braun & Clarke, 2006; Vaismoradi *et al.*, 2013). We invite the reader to cautiously consider the synthesis of qualitative results as an oversimplification of the students' rich discourses, often ambivalent.

5. Results

Overall, 49 students took part in the study. However, 1 student did not answer the final survey, and 2 essays from 2 students had to be removed

for being incomplete. Therefore, the final number of cases where 48 for the surveys and 47 for the essays. Moreover, our study was ideographic and interpretive. The data gathered from the survey and the essays were different and treated as two separated samples of data focused on the engagement with the instruments rather than on the individual response (no association intra-subject between the two moments of data-collection).

5.1. *Standard survey about SET*

As a first step, we analyzed the survey on students teaching quality evaluation (Ghislandi *et al.*, 2019). The calculations are based on a sample of N = 48. Over 12 questions, for 7 questions the positive answers («Rather Yes» and «Definitely Yes») exceed the negative ones, with positive aggregated answers equal or superior to 66.67% (see *Tab. 1* for these 7 questions: Q01, Q04, Q05, Q08, Q09, Q10, Q11). The students, moreover, think that the material provided is adequate for the study (Q03: 60% positive aggregated answers). However, there are negative aggregated results above the 50% for 4 questions. The students expressed their dissatisfaction with: the workload (Q02: 75% of negative aggregated results); the teacher's ability to stimulate the students' interest in the course's topic (Q06: 52%); the clarity of teachers' explanations on the topics covered by the course (Q07: 52%). We also found a slight prevalence of negative responses (Q12: 54%) when looking at students' overall satisfaction with the course.

The quantitative data presented a description of the situation, but did not explain the causes that triggered the not completely positive reactions regarding overall satisfaction. Moreover, in the standard survey, the space for open answers led to gather very few answers, from those less satisfied with the course (2 comments over 48 answers; 129 words against the 60,859 words gathered in the essays). A plausible interpretation is related to the course workload which had the highest percentage of dissatisfaction. Having introduced a high number of resources and activities might have made almost half of the students feel that the teacher did not cover all topics clearly and did not provide appropriate support. However, this hypothesis could not be fully explored throughout the data yielded by the survey.

Table 1 contains the questionnaire adopted and *Figure 1* the statistics commented above.

Table 1. – Students' standard survey about SET.
Questions translated from the original Italian version.

Question Number	Question Text
Q01	Was the preliminary knowledge possessed sufficient to understand the topics included in the exam program?
Q02	Is the students' workload proportionate to the credits assigned?
Q03	Is the teaching material (indicated and/or provided) adequate for the study of the subject?
Q04	Were the methods of examination clearly defined?
Q05	Are the course schedule, exercises and other educational activities respected?
Q06	Does the teacher stimulate/motivate the interest in the discipline?
Q07	Does the teacher explain the topics clearly?
Q08	Are the integrative teaching activities (tutorials, laboratories, etc.), where existing, useful to learn the subject?
Q09	Has the teaching been carried out in a manner consistent with what was stated on the course website?
Q10	Is the teacher available for clarifications and explanations?
Q11	Are you interested in the topics covered in the course?
Q12	Overall, are you satisfied with how the course went?

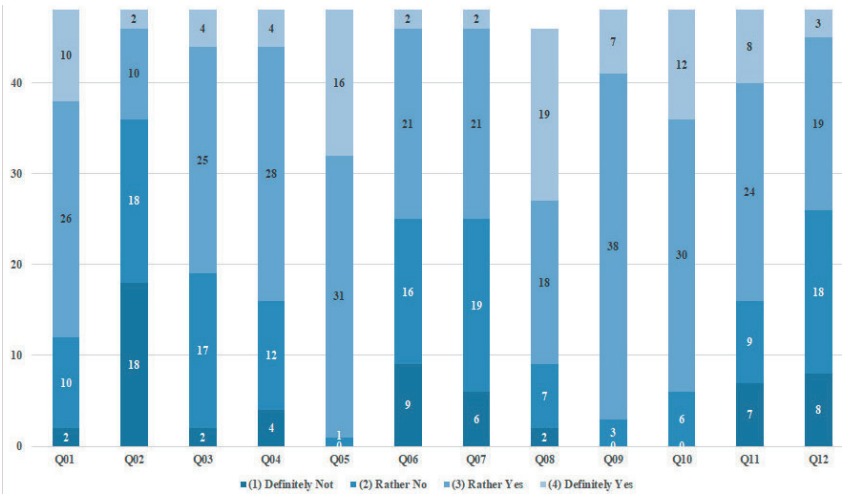

Figure 1. – Descriptive statistics from the students' standard survey about SET.

5.2. *Final essay*

The qualitative analysis of the final essay involved 49 written works, from which 2 had to be removed for being incomplete. The corpus of 60,859 words (about 650 pages) was extracted and codified through a process of thematic analysis, yielding 1171 codes (for more details on the codebook containing the codes, categories and exemplar excerpts of discourse, see Ghislandi *et al.*, 2019). An exemplar text bracketed out from the corpus is also presented to make the procedure more transparent in *Table 2*.

Table 2. – Example extracted from the full codebook, qualitative analysis (Ghislandi et al., 2019).

Label	Codes	Nr. of Codes	Exemplar Bracketed Text
Course (-)	Course (-) [Negative impressions]	49	WkD- [Gir4-P9] *I find the overall learning design is excellent for this course; however, I believe that the students' workload is maybe too high, in relation to the credits released by the end of the course.*
WkD-	Workload and deadlines	34	
WeLD-	Whole course eLD negative	6	
T/Scom-	Teacher/Student communication	3	
COrg-	Content organization	2	
SevDifficulties	Several difficulties experienced along the course	4	

The inductive process of codification led to identify a range of neutral, positive and negative impressions on the course and the teaching methods adopted. *Figure 2* summarizes the codes and shows the frequency of the main themes, where less numerous neutral expressions against the positive and negative students' expressions are highlighted and aggregated.

There is a relationship of 118(+) and 38(-) for the evaluation of collaborative learning; of 181(+) and 4(-) for the students' appreciation on self and peer evaluation; of 123(+) against 9(-) on rubrics supporting self and peer evaluation. The overall teaching methods adopted in the course were expressed with 207(+) and 14(-). Only positive expressions (22+) were found for the metacognition. However, the students found the technological tools harder to evaluate, since there was more dispersion in their opinion, with 77(+) and 62(-). We observed, moreover, that 160 text units relate

to positive expressions about the whole course, without further comments on a specific aspect, against 49 negative expressions. All in all, there were 888/1171 positive expressions (74%), including teaching methods, tools, learning activities, etc., against 176/1171 negative expressions (15%).

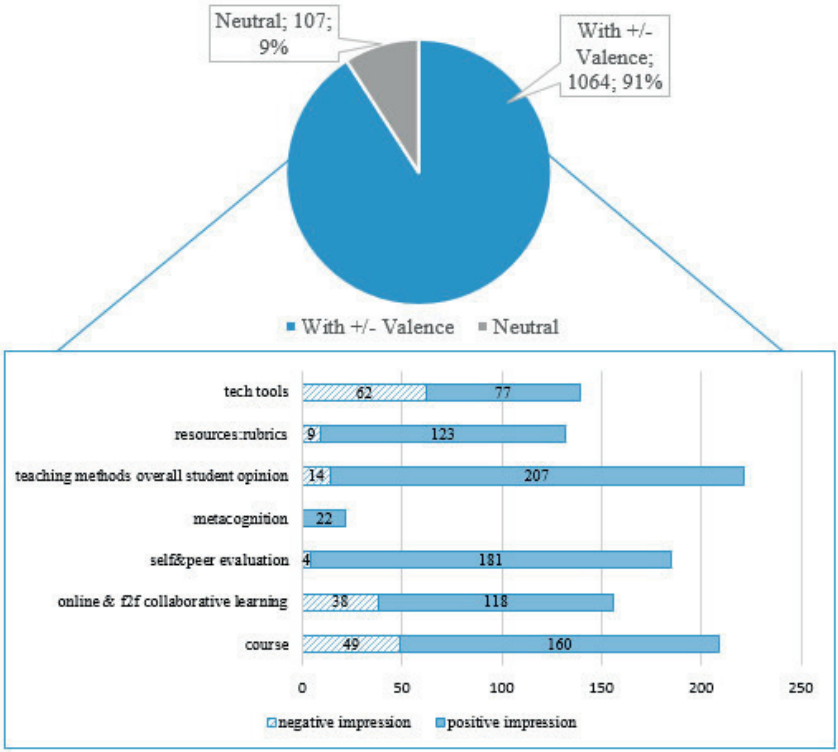

Figure 2. – Valence assigned by the students (Neutral, Positive and Negative Impressions about the eLD course in the students' final essays) using the codes identified in the qualitative analysis.

These results support the idea that the course encompassed a rich and lived experience for most students. Among the instructional strategies designed for the course, the ones perceived as more innovative and above all effective were the authentic tasks (eLearning design of a module to be used by the tandem group; the self-evaluation of one's own work by the design and delivery rubrics; the peer evaluation by the tandem group, etc.). Many students have also appreciated the possibility of reflecting on their learning through

the final essay. The rather controversial opinion on technological tools, instead, shows a level of difficulty and the alleged conflict that may have arisen when the students had to deal with the technologies of the course, that is personal computer and forum instead of mobiles and instant messaging, that represent the students' daily experience with digital technologies.

The qualitative analysis points out at an existing culture of learning where active methods and intensive use of digital environments could be more an exception than the rule, and to the complexity of two contrasting learning cultures (traditional, face-to-face and more directive teaching against active methods).

5.3. *Spotting differences and convergences between methods*

In this paragraph, we undertake the data triangulation process, which is based on an interpretative effort to spot the differences and convergences of the two instruments under analysis (quantitative survey and qualitative essays) in capturing the perceived quality of teaching as evaluated by the students. In this regard, it is important to consider that the triangulation involves the collation and comparison of data from multiple sources at multiple levels (Babones, 2016).

The quantitative survey focused on the synthesis of emerging issues relating to the teaching methods, but demonstrated to be less sensitive – at least in its standard form used nowadays in Italy – to the deeper phenomena entangled in the socio-constructivist course under evaluation. The qualitative analysis, instead, was less effective for summarizing the issues, with the students expressing their opinions in ambiguous or even conflicting ways. However, in this case the students' discourse showed a plurality and richness of perspectives that allowed the emergence of their judgment on innovative teaching strategies and tools adopted in the course. While the workload problem was evident in both evaluation approaches, the students' survey and the final essay, the latter enabled the reader to understand that a few students, rather than being simply dissatisfied, reflected on the fact that the more active teaching method required more work than a traditional course.

6. Discussion

The most relevant result of our study is given by the substantial diversity of the issues that can be captured when the SET measurement is conducted through the questionnaire if compared with the qualitative method that we

have applied to the final essay produced by the students. In our case study, we found that the two methods capture different phenomena, since they are, above all, developed from different theoretical frameworks (Crotty, 1998). In fact, as argued above, in the questions of the Italian questionnaire there seems to be an implicit assumption on a standardized approach to teaching. However, the course analyzed in the paper has been developed from a constructivist approach which can be deemed as an emergent practice adjusting to alternative pedagogical frameworks that are well grounded in the international literature (Jonassen *et al.*, 1995; Hmelo-Silver *et al.*, 2007). Accordingly, the experience of the students in this type of courses concerns aspects of active participation, reflection and development of soft skills. These aspects were detected through the qualitative analysis but were invisible in the survey data. In other words, the analysis of the essays provides preliminary data confirming that the existing survey overlooks important aspects of the students' experience that are crucial for the constructivist pedagogical approach. Further research is needed to examine if and how this problem can at least partially be solved by revising some items of the existing survey. However, we guess that triangulating the results of the survey with a phenomenological analysis of students' experience would be beneficial in this sense (Lattuca & Domagal-Goldman, 2007).

The rationale for this guess is that the way in which the system is deployed reveals the theoretical framework on which the evaluation is based. Misalignments between the theoretical approach of the evaluation and of the pedagogical design and practice, could end up «obscuring» positive/innovative teaching strategies. It is not about banning traditional ways of teaching from higher education, since it can certainly be effective when it is used in the context of a «design for learning» (Laurillard, 2012) that is based on a sophisticated and refined theoretical perspective (Ghislandi, 2005). However, in order to effectively embrace Quality Enhancement as a practice of continuing improvement, our research suggests that theoretical approaches, methods and instruments have to be congruent. This is consistent with the issue of the methodological coherence reported by Richards *et al.* (2009) in qualitative research, between research question, methodologies and methods. Our research suggests this congruence should be kept also in *evaluation* methods, particularly applied to emerging socio-constructivist teaching practices.

The results we have exposed are only preliminary and present some limits. First of all, our analysis is a case study and it is targeted to the Italian SET, particularly for what the quantitative analysis is concerned. Moreover, the main weakness is that the students knew that their non-anonymous final essays would be read by the teacher before they would

receive the final grade. This certainly may have led some of them to try to please the teacher, rather than assess the characteristics of the course in a sincere manner. To ensure greater validity to the analysis, in the future course we will design the meta-reflective part of the final essay separately from the evaluation of the course, so that this latter can be anonymous.

7. Conclusion

The study gave results from the practical as well as theoretical and political point of view. Concerning the practical suggestions, the survey as well as the analysis of the final essay brought many tips for improving the design of the course and optimizing the teaching approach.

Concerning the theoretical and political aspects, the study reminds us that the quality of an object or a process is very much linked to stakeholders, to the epistemologies theoretically defining the quality, to the methodologies characterizing it, to the instruments adopted for the measurement, and to the students' expectations.

The standard questionnaire should be integrated, for a more correct analysis of the teaching quality, with other questions and with other methodologies and tools that leave room for more in-depth analysis of the students' opinions. The issue is that the analysis of qualitative data requires more resources than the analysis of the questionnaire data, which can be processed semi-automatically.

Finally, in the courses time have to be allocated to overcome the impression, reported by Spencer and Schmelkin (2002), that students have little faith that teachers pay attention to SET results. Promoting this aspect can make the students more conscious of SET and its results, facilitating stronger engagement and positivity towards the activities carried out in their institution.

Aknowledgments

We would like to sincerely thank Elisabeth Julie Vargo and Stefano Bannò for the revision of the English text.

The research is partially funded by Laboratorio Interdisciplinare per la Qualità e l'Innovazione della Didattica (LIQuID), Piano Strategico d'Ateneo 2017-21, Università degli Studi di Trento.

References

Allendoerfer, C., Wilson, D., Plett, M., Bates, R. A., Smith, T. F., & Veilleux, N. M. (2016). Student perceptions of faculty support: Do class size or institution type matter? In *ASEE Annual Conference and Exposition, Conference Proceedings* (June).

ANVUR (2018). *Linee di indirizzo per lo sviluppo professionale del docente e strategie di valutazione della didattica in Università*. Roma.

Babones, S. (2016). Interpretive quantitative methods for the social sciences. *Sociology*, *50*(3), 453-469. https://doi.org/10.1177/0038038515583637

Balzaretti, N., & Vannini, I. (2018). Promuovere la qualità della didattica universitaria. La Formative Educational Evaluation in uno studio pilota dell'Ateneo bolognese. *Journal of Educational, Cultural and Psychological Studies*, *18*, 187-214. https://doi.org/http://dx.doi.org/10.7358/ecps-2018-018-balz

Bernardi, L., Boncori, L., Gola, M., Magni, C., & Ottaviani, M. G. (1998). *Valutazione della didattica da parte degli studenti. Rapporto finale del gruppo di ricerca*. Roma.

Bertaccini, B. (2015). Il sistema italiano di valutazione della didattica universitaria. Analisi critica della normativa vigente / The Italian approach in evaluating the quality of the academic teaching system: A critical review of the current regulation. *RIV – Rassegna Italiana di Valutazione*, *63*, 7-22. https://doi.org/10.3280/RIV2015-063002

Bertaccini, B., Giusti, A., & Petrucci, A. (2019). Students' opinions on teaching and services provided by the Italian Universities: A proposal for a new evaluation scheme. *Quality & Quantity*, *53*, 1037-1050. https://doi.org/10.1007/s11135-018-0801-y

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71-88.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Buck, D. (1998). Student evaluations of teaching: Measure the intervention, not the effect. *American Psychologist*, 1224-1226.

Cacciamani, S., & Messina, R. (2011). Knowledge Building Community. Genesi e sviluppo del modello. *Qwerty-Open and Interdisciplinary Journal of Technology, Culture and Education*, *6*(2), 32-54.

Chiandotto, B., Grilli, L., & Rampichini, C. (2005). *Valutazione dei processi formativi di terzo livello. Contributi metodologici*. Firenze: Università degli Studi di Firenze.

Cole, K. A., Barker, L. R., Kolodner, K., Williamson, P., Wright, S. M., & Kern, D. E. (2004). Faculty development in teaching skills: An intensive longitudinal model. *Academic Medicine: Journal of the Association of American Medical Colleges*, *79*(5), 469-480.

Communique, Y. (2015). *European Higher Education Area (EHEA)*.

Creswell, J. W., & Plano Clark, V. L. (2007). Choosing a mixed methods design. In J. W. Creswell & V. L. Plano Clark (Eds.), *Designing and conducting mixed methods research* (1st ed., pp. 58-88). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. London: Sage.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. London: Sage.

Ehlers, U. D. (2004). Quality in e-learning from a learner's perspective. *European Journal of Open, Distance and E-learning*, *7*(1).

Elassy, N. (2015). The concepts of quality, quality assurance and quality enhancement. *Quality Assurance in Education*, *23*(3), 250-261. https://doi.org/10.1108/QAE-11-2012-0046

Fabbris, L. (2002). La misura della student satisfaction per la valutazione della qualità della didattica. In F. Delvecchio & L. Carli Sardi, *Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari* (pp. 1-20). Padova: Cleup.

Falchikov, N. (2005). *Improving assessment through student involvement*. London: Routledge-Falmer.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906.

Ghislandi, P. (2012). adAstra: A rubrics' set for quality eLearning design. In P. Ghislandi (Ed.), *eLearning: Theories, Design, Software and Applications* (pp. 91-106). Rijeka: InTech - Open Access Publisher. https://www.academia.edu/26456212/adAstra_A_Rubrics_Set_for_Quality_eLearning_Design

Ghislandi, P., Raffaghelli, J., & Cumer, F. (2012). La qualità dell'eLearning. Un approccio qualitativo per l'analisi dei feedback degli studenti e dei docenti. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, *7*(2), 25-47. https://doi.org/10.6092/ISSN.1970-2221/3218

Ghislandi, P., Raffaghelli, J., & Yang, N. (2013). Mediated quality: An approach for the eLearning quality in higher education. *International Journal of Digital Literacy and Digital Competence (IJDLDC)*, *4*(1), 56-63.

Ghislandi, P. M. M. (2005). *Didattiche per l'università*. Trento: Università degli Studi di Trento.

Ghislandi, P. M. M., Raffaghelli, J. E., Sangrà, A., & Ritella, G. (2019). Appendices – The street lamp paradox: Analysing students' evaluation of teaching through qualitative and quantitative approaches. https://doi.org/10.5281/ZENODO.3560223

Gibbs, G. (2010). *Dimension of qualaity in higher education*. York, UK: Higher Education Academy.

Giovannini, M. L. (1988). *La valutazione delle innovazioni nella scuola*. Bologna: Cappelli.

Giovannini, M. L., & Silva, L. (2014). Le ricerche sui questionari-studenti per la valutazione dell'insegnamento universitario. Quali elementi di problematicità in rapporto all'uso delle risposte? *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, *9*(3), 19-51.

Grion, V., & Tino, C. (2018). Verso una «valutazione sostenibile» all'università. Percezioni di efficacia dei processi di dare e ricevere feedback fra pari. *Lifelong Lifewide Learning*, *14*(31), 38-55.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Harvey, L. (2004). *Analytic quality glossary, quality research international*. http://www.qualityresearchinternational.com/glossary/

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9-34. https://doi.org/10.1080/0260293930180102

Henard, F., & Roseveare, D. (2012). Fostering quality teaching in higher education: Policies and practices. *An IMHE Guide for Higher Education Institutions*, 7-11.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *American Journal of Distance Education*, *9*(2), 7-26.

Jonassen, D. H. (1992). Evaluating constructivistic learning. In M. T. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction* (pp. 137-148). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kaplan, A. (1973). *The conduct of inquiry*. Transaction Publishers.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495-523.

Lalla, M. (2006). *Problemi e osservazioni sul questionario utilizzato per la valutazione della didattica da parte degli studenti frequentanti*. Working Paper. Modena.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174.

Lattuca, L. R., & Domagal-Goldman, J. M. (2007). Using qualitative methods to assess teaching effectiveness. *New Directions for Institutional Researches*, Special Issue: *Using Qualitative Methods in Institutional Assessment*, *136*, 81-93. https://doi.org/10.1002/ir.233

Laurillard, D. (1993). *Rethinking university teaching: A framework for the effective use of educational technology*. London: Routledge.

Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. London: Routledge.

Lincoln, Y., Lynham, S., & Guba, N. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (4th ed., pp. 97-128). Thousand Oaks, CA: Sage.

Loperfido, F. F., Ligorio, M. B., & Cole, M. (2011). Blended approach per la costruzione collaborativa e partecipativa. *Qwerty-Open and Interdisciplinary Journal of Technology, Culture and Education*, *6*(2), 274-287.

Losito, B. (1996). Valutare per capire: la valutazione a sostegno dei progetti di innovazione. *Italian Journal of Educational Technology*, *4*(1), 43-63.

Macdonald, R. (2006). The use of evaluation to improve practice in learning and teaching. *Innovations in Education and Teaching International*, *43*(1), 3-13. https://doi.org/10.1080/14703290500472087

Mahoney, C. (2012). How should teaching, learning and assessment evolve? How to drive quality teaching. *Blue Skies: New Thinking about the Future of Higher Education*.

Muukkonen, H., Lakkala, M., & Paavola, S. (2011). Promoting knowledge creation and object-oriented inquiry in university courses. *Learning across Sites: New Tools, Infrastructures and Practices*, 172-189.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102-122. https://doi.org/10.1080/02602938.2013.795518

Ory, J. C. (2001). Faculty thoughts and concerns about student ratings. *New Directions for Teaching and Learning*, *87*, 3-15. https://doi.org/10.1002/tl.23.

Pagani, L., & Seghieri, C. (2002). A statistical analysis of teaching effectiveness from students' point of view. *Developments in Statistics*, *17*, 197-208.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Raban, C. (2007). Assurance versus enhancement: Less is more? *Journal of Further and Higher Education*, *31*(1), 77-85. https://doi.org/10.1080/03098770601167948

Rampichini, C., Grilli, L., & Petrucci, A. (2004). Analysis of university course evaluations: From descriptive measures to multilevel models. *Statistical methods and Applications*, *13*(3), 357-373.

Richards, L., Morse, J. M., Gatti, F., & Graffigna, G. (2009). *Fare ricerca qualitativa. Prima guida*. FrancoAngeli.

Rienties, B. (2014). Understanding academics' resistance towards (online) students evaluation. *Assessment and Evaluation in Higher Education*, *39*(8), 987-1001. https://doi.org/10.1080/02602938.2014.880777

Ritella, G., Di Maso, R., Mclay, K., Annesse, S., & Ligorio, M. B. (2020). Remembering, reflecting, reframing: Examining students' long-term perceptions of an innovative model for university teaching. *Frontiers in Psychology*. Advance online publication.

Sangrà, A., Vlachopoulos, D., & Cabrera, N. (2012). Building an inclusive definition of e-learning: An approach to the conceptual framework. *The International Review of Research in Open and Distributed Learning*, *13*(2), 145-159.

Shaw, I. (1999). *Qualitative evaluation*. London: Sage. https://doi.org/10.4135/9781849209618

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspective on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, *27*(5), 397-409.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642. https://doi.org/10.3102/0034654313496870

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249-276. https://doi.org/10.3102/00346543068003249

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22-42.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398-405. https://doi.org/10.1111/nhs.12048

Wegerif, R. (2006). A dialogic understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning*, *1*(1), 143-157.

Yang, Nan (2015) *Quality teaching in large university classes: Designing online collaboration among learners for deep understanding.* PhD Thesis, University of Trento.

Riassunto

In questo articolo si riporta uno studio sulla valutazione dell'insegnamento accademico svolta dagli studenti, o Student Evaluation of Teaching (SET). Considerando che la qualità dell'insegnamento accademico può essere esaminata a vari livelli, la sua valutazione

può richiedere una varietà di metodi. Quello più comune, utilizzato abitualmente nelle università italiane, consiste nella compilazione di un questionario alla fine del corso. In questo studio, effettuato su un insegnamento universitario offerto nell'anno accademico 2016/2017, si propone l'analisi qualitativa dell'elaborato di fine corso che riguarda il percorso di apprendimento e la riflessione degli studenti sulla qualità dell'insegnamento. L'analisi qualitativa degli elaborati è stata poi comparata con i risultati del questionario SET. 48 studenti hanno compilato il questionario, e 47 hanno anche consegnato l'elaborato (650 pagine di testo in totale). La ricerca mostra che mentre il questionario fornisce informazioni utili per la Quality Assurance (QA) a livello istituzionale, l'analisi qualitativa degli elaborati ha permesso di comprendere la reazione degli studenti ai diversi elementi specifici dell'approccio pedagogico adottato. Il nostro studio suggerisce che può essere utile, almeno per i corsi basati su un paradigma di apprendimento socio-costruttivista, considerare l'integrazione del tradizionale questionario SET con metodi alternativi per la valutazione dell'insegnamento, che permettano di raccogliere informazioni qualitative, non rilevate dal questionario, ma utili per la Quality Enhancement (QE) dell'insegnamento.

Parole chiave: Apprendimento socio-costruttivista; Quality Assurance; Quality Enhancement; Student Evaluation of Teaching; Valutazione qualitativa della didattica.

How to cite this Paper: Ghislandi, P., Raffaghelli, J., Sangrà, A., & Ritella, G. (2020). The street lamp paradox: Analysing students' evaluation of teaching through qualitative and quantitative approaches [Il paradosso del lampione: analizzare, attraverso approcci qualitativi e quantitativi, la valutazione di un insegnamento accademico da parte degli studenti]. *Journal of Educational, Cultural and Psychological Studies*, *20*, 65-85. DOI: https://dx.doi.org/10.7358/ecps-2020-021-ghis